# Web-DBサーバと連携する音声ラベリングプログラムの開発

# Web-DBサーバと連携する音声ラベリングプログラムの開発

小杉　風友・今野　英明・金光　秀雄

北海道教育大学函館校情報基礎研究室

# Development of Speech Labeling Program Communicating with Web-Database Server

KOSUGI Kazetomo, KONNO Hideaki, and KANEMITSU Hideo

Department of Information Sciences, Hakodate Campus, Hokkaido University of Education, Hakodate, 040-8567

## 概　要

音声データベースを一元化して複数の利用者で共有し，ネットワークを介して利用できるようにシステム化すれば，音声データの収集や管理および配布にかかるコストの削減が期待できる。一方，著者らはWebデータベース汎用システムを開発して活用している。このシステムに音声データと，それに付随する種々のデータを関連付けて格納することにより，音声データベースシステムとしての利用が可能になる。そのために，本研究ではシステムに格納された音声データを取得し，GUIによる操作で音素ラベリングを行うためのプログラムを開発した。また，このプログラムが音素ラベリングの結果を即時にシステムに送信し，それを元の音声データと関連付けてサーバに格納できることを動作試験により確認した。

## １．はじめに

音声認識や音声合成を始めとする今日の音声処理では，確率モデルの構築や統計的処理のために大量の音声データが必要である。音声の音響的特徴を調査することを目的とする音声分析においても多くのデータは欠かせない。このような大量の音声データを音声データベースとして一元化して複数の利用者で共有し，ネットワークを介して利用できるようにすれば，音声データの収集や管理および配付にかかるコストの削減が期待できる。

また，一般に音声処理では，音声波形のデータに加え，発話の開始・終了時刻や音素位置のデータ（音素ラベル），話者の情報等が必要となることが多い。そのため，音声データベースでは，音声波形のデータからその音声に関わる各種データを作成し，提供する必要がある。

青木らは音声合成用の音声データベースをオープンコンテンツ化することを目的とした研究を行っている[1]。この研究では音声のラベルや各種情報をXML（Extensible Markup Language）で記述し，それをWAV（RIFF）形式のファイル内に格納した。このWAVファイルは通常の音声ファイルとして再生処理等が可能であり，さらに単一のファイル内に音声波形のデータと付加情報が入っているためデータの管理効率が良い。ただ

し，WAVファイルはバイナリファイルであるため，XMLデータの記述と閲覧には専用のエディタが必要となり，データの汎用性は低下する。

著者らはWebデータベース汎用システムを開発し，授業での資料提示やオンライン試験，研究室内向け文書の保存等に利用してきた。このシステムは，Webページを通じて種々のデータを関係データベースに格納し，利用者に提示するものであり，以下の特徴がある。

(a) 自動的に生成されるIDでデータを識別するため，データの登録時にデータ名の重複を意識する必要がない。このため，通常のファイルに比較してデータ管理が容易である。

(b) ユーザ認証機能を利用して，データの登録や取得に関する種々の権限を設定できることから，不特定利用者へのデータ公開の用途に加え，特定の利用者間でのデータの共有にも適している。

(c) データおよびそのデータから派生したデータを，データ間の関係と共にシステムに格納できる。

従って，このシステムに音声データと，それに付随する種々の情報や分析結果等を格納すれば，特定の利用者間で音声データベースを共有するためのシステムとして利用できる[2]。

このシステムの操作は基本的にWebブラウザのみで行えるものの，データ処理の用途には，処理を行う外部プログラムがデータを取得し，その処理結果をシステムのサーバに送信できる必要がある。そのためのアプローチとして次の二つが考えられる。

1）システムと直接に通信する外部プログラムを自ら作成する[3,4]。

2）外部プログラムとシステムとの間のインターフェースプログラムを用意する。外部プログラムはインターフェースプログラムを通じて，システムとデータの受け渡しをする[5]。

前者のアプローチ1）の利点は，作成者がプログラムの詳細を把握しているため，さまざまな局面に柔軟に対応でき，容易に改良ができることである。ただし，プログラムの作成には時間と手間がかかる。後者のアプローチ2）では，既存の外部プログラムをそのままで，あるいは若干の変更だけで利用できることが利点である。一方で，変更が不可能な外部プログラムの場合，インターフェースプログラムとの通信ができず，システムを利用できない可能性が生じる。

本稿では，アプローチ1）に基づき，Webデータベース汎用システムのサーバから音声データを取得し，それを基に作成したデータを即時に元のデータと関連付けてサーバに格納するプログラムについて述べる。特に，音声波形のデータを基に音素ラベルを作成する処理（音素ラベリング）はある程度の自動化が可能であるものの，正確なラベリングは人手による作業を要する手間のかかる処理である。そこで，このラベリング処理を主として行うGUI（Graphical User Interface）を持つアプリケーションプログラム（以降“GuiAudio”と呼ぶ）を作成したので，これについて報告する。

## 2．Webデータベース汎用システムの動作概要

本研究で作成したアプリケーションプログラムGuiAudioはWebデータベース汎用システムにおけるクライアントとして動作する。クライアントを含むWebデータベース汎用システムの構成は図1のとおりである。このシステムでは，クライアントがHTTPプロトコルでリクエストをHTTPサーバに送る。サーバは指定されたURL

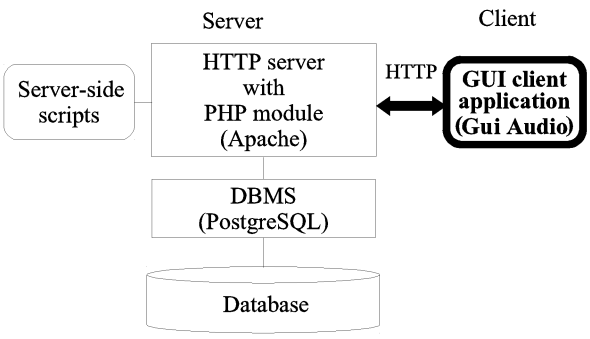


**図1**　Webデータベース汎用システムの構成

に応じてサーバサイドスクリプトを実行し，データの読み書きをDBMS（Database Management System）に要求する。DBMSはデータベース内のデータをHTTPサーバに渡し，これを使って動的に生成されたWebページをHTTPサーバがクライアントに送信する。

## 3．GuiAudioの機能

本研究で開発したGuiAudioは次の(a)から(e)の機能を持つ。

(a)　ユーザ認証への対応機能

サーバにデータを送信する場合にユーザ認証が必要となる。GuiAudioではchallenge-response方式のユーザ認証に対応しており，そのためのログインIDやパスワードを送信できる。

(b)　簡易ブラウザ機能

GuiAudioはサーバから送信されるHTMLを解釈して表示する。この簡易ブラウザ機能は処理する音声データをサーバから選択する際に使われる。

(c)　音声データの再生・波形表示・スペクトログラム表示機能

簡易ブラウザ機能を使ってダウンロードした音声データの再生や波形表示，スペクトログラム表示を行う。これらの機能はラベリングの操作において必要となる。

(d)　ラベリング機能

(c)で示した機能を利用して，プログラムの利用者がラベリングを行える。音素などの入力にはキーボードを利用し，その音素の開始時刻と終了時刻の確定はマウスを利用する。

(e)　データの送信機能

(d)までの機能を利用して行ったラベリングの結果を格納したファイルや，音声データの諸情報を格納したファイルなどをWebデータベース汎用システムのサーバに送信する。このとき音声データと各種データを関連付けるための情報も送信される。

## 4．開発言語

本研究では，標準ライブラリの豊富さ等の理由からJava言語を用いることとした。Java言語によるプログラミングでは，Javaアプリケーションとアプレットの形態が考えられる。

西村らはJavaアプレットをWebページに埋め込むことによって音声入力用のGUIインターフェースを実装し，Webを使った音声データの収集システムを作成することに成功している[6]。また，青木らはサーバから送信する音声データをWebブラウザで再生するためにJavaアプレットを用いている[7]。このようにJavaアプレットにはWebページに組み込んでクライアントのWebブラウザ上で音声を扱える利点がある半面，セキュリティ上の理由からJavaアプレットによるローカルファイルへのアクセスは厳しく制限されている[8]。このためクライアントで行った処理結果をファイルに保存する場合，処理結果をサーバに送信することが困難になる。そこで，本研究では，プログラムによるローカルファイルへのアクセスの自由度を優先し，クライアントで通常の方法でプログラムを起動させるJavaアプリケーションとしてプログラムを開発することとした。

## 5．GuiAudioの構成

GuiAudioの構成を図2に示す。GuiAudioはMainモジュールが各種モジュールを呼び出し，その結果を受け取って必要な機能を利用者に提供する。

PreAccessモジュールでは事前に必要なログインなどの処理を行い，セッションを継続するためのクッキーを保持する。DownloadDataモジュールではクッキーを使ってサーバから音声データをダウンロードする。

AudioPlayerモジュールは音声再生機能を提供する。音声データの再生は任意の時間から可能である。AudioAnalyzeモジュールでは，音声データからサンプリング周波数や量子化ビット数など

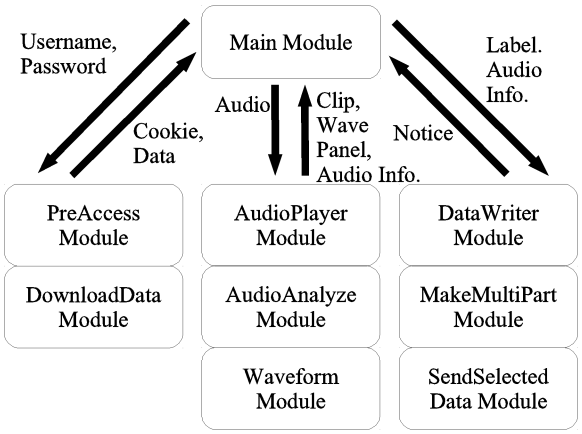


**図２**　クライアントプログラムの構成

の諸情報の抽出し，音声波形の振幅値の復号化やFFT 演算を行う。 Waveform モジュールではAudioAnalyze モジュールで取り出した音声データの振幅値を元にして音声波形を描画するための数値列を生成する。

DataWriter モジュールでは，サーバに送信するデータをファイルとしてクライアントの単一のディレクトリに保存する。 MakeMultiPart モジュールでは，書き出したファイルをサーバに送信するために必要となる HTTP リクエストボディ内のマルチパート部を作成する。 SendSelectedData モジュールでは，マルチパートに HTTP リクエストヘッダを付加して HTTP リクエストを完成させ，サーバに送信する。元の音声データと送信データを関連付けるための情報も送信する。

## ６．実行過程

GuiAudio の開発では，限られたウィンドウスペースを有効に使うため，操作のステップ別にタブを用意した。利用者は STEP1: Select Audio Data， STEP2: Labeling， STEP3: Send Result の順にタブ内で操作を行う。各タブにおける操作が完了すると次のタブへ自動で遷移する。

まず， GuiAudio を起動して ID と Password を入力する。次にログイン先を選択して Login ボタンを押すと，サーバへログインを試みる。ログインが正常に完了すると，サーバから送信される HTML データを受け取り，図３に示すとおり，タブ STEP1内に Web ページが出力される。

利用者が処理対象となる音声データを Web ページより選択すると，タブは STEP2に遷移し，ラベリングのための GUI を提供する。図４に単語/aisatsu/を選択した後のタブ STEP2の画面を示した。波形パネル上の任意の位置にマウスポインタを置き，右クリックすると音声データをその位置から再生できる。ここでは利用者が，再生機能や波形，サウンドスペクトログラムを手がかりに音素ラベリングを行う。波形パネル上では１度目の左クリックで音素の開始時刻が入力され， ２度目の左クリックで終了時刻が入力される。音素の開始時刻と終了時刻を確定した後，音素をテキストフィールドに入力して Set ボタンを押すと，右側のテキストエリアに音素とその開始時刻，終了時刻が入力される。この一連の操作を繰り返してラベリングを行う。 Complete ボタンを押すとテキストエリア内のデータが保存され，送信の際に利用される。また，画面がタブ STEP3に遷移する。

タブ STEP3では送信の操作を行う（図５）。まず，送信したい項目をチェックボックスで選択する。 Phoneme 欄は利用者が入力する音素列，Comment 欄は話者の情報などが入力されたコメント， AudioInfo は音声データのサンプリング周波数や量子化 bit 数，バイト順などの諸情報，Amplitude は振幅値列， LabelData はラベリングの結果， FFT は FFT 演算の結果である。Send ボタンを押すとチェックした項目のデータと共に，音声データと各種データを関連付けるための情報がサーバへ送信される。サーバではこの関連付け情報を元に，音声データと各種データを関連付けて格納する。

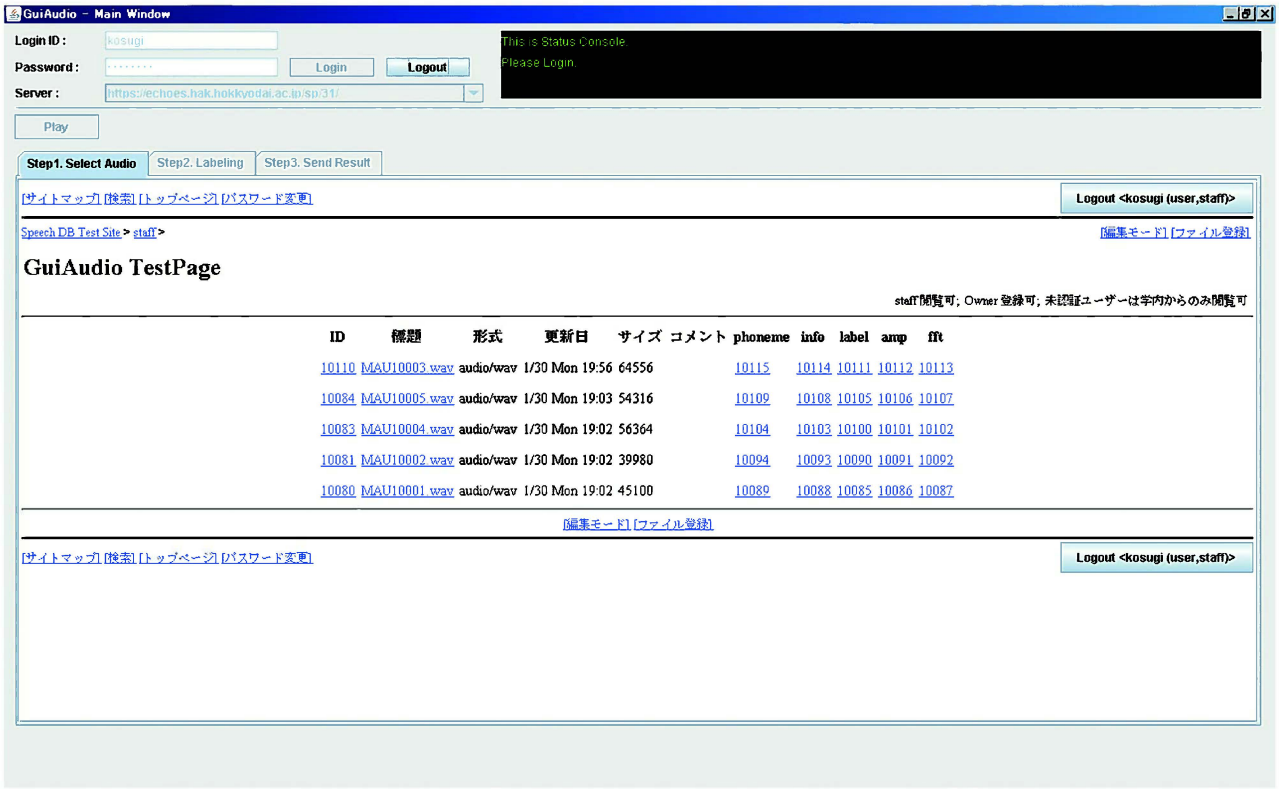


**図 3**　音声データの選択画面

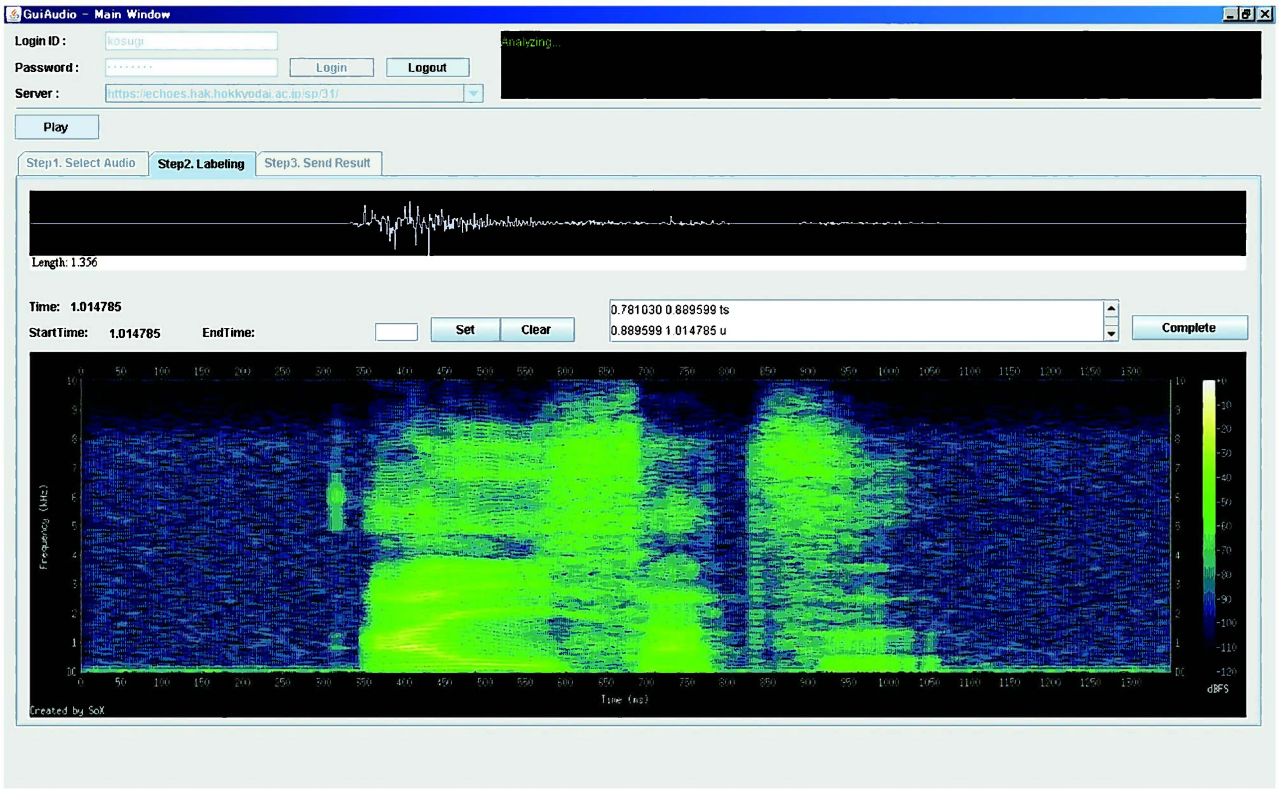


**図 4**　音声の再生・表示・ラベリング画面

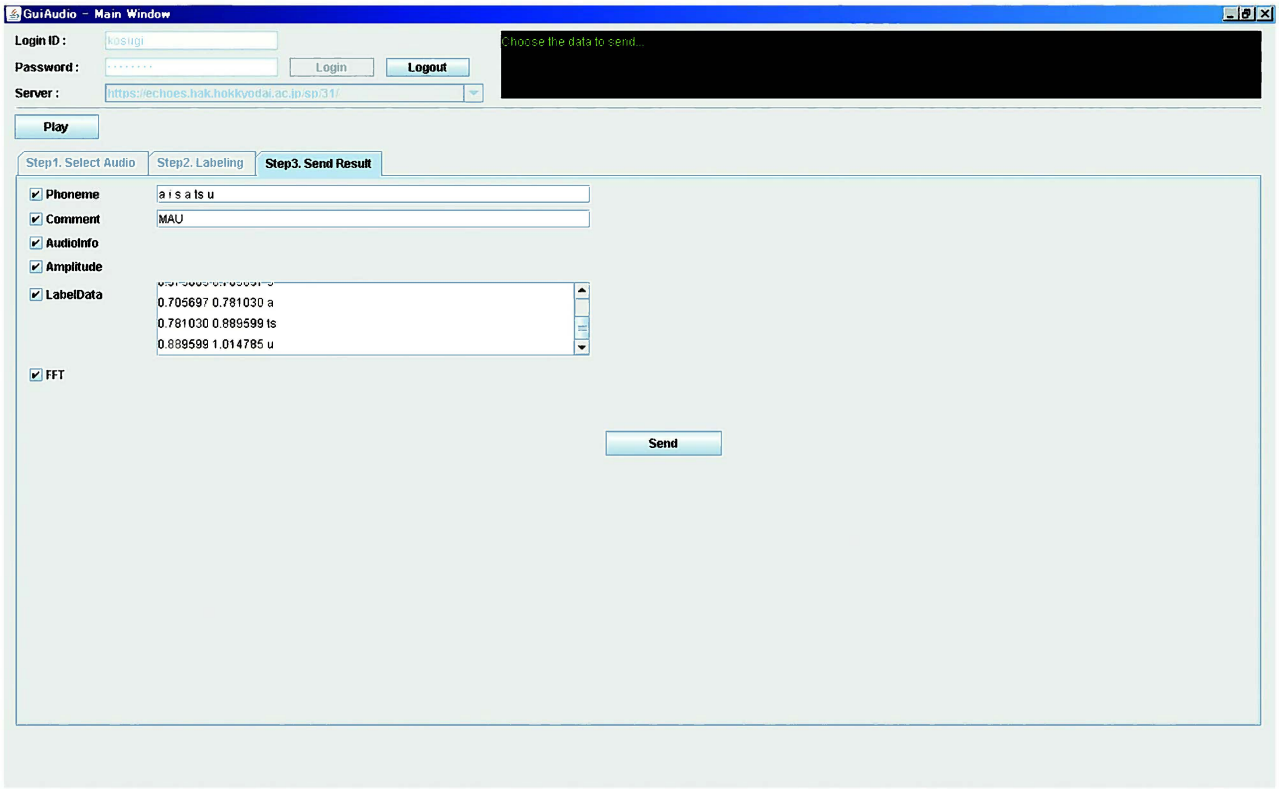


**図 5**　分析結果とラベリング結果の送信画面


**図 6**　データ送信後の web ページ

## 7．送信結果の確認

データが正しくサーバに送信されたことを確認するために，図 3 に示した Web ページを Web ブラウザで表示したものが図 6 である。

図 6 において， phoneme, info, label, amp, fft の各列に記述された数値は， ID 列と標題列に示された元の WAV 形式の音声データに関連付けられたデータの ID であり，図 5 においてサーバに送信したものである。 Web ブラウザ上でクリックすることで，その内容を確認することができる。

これらの関連付けの情報は Web データベース汎用システムのサーバ内では，関係データベースのテーブルに格納される。表 1 に ID が10080の音声データに関する行のみを例示する。このテーブルではデータに付与された ID を src_id 列と dest_id 列に格納し， link 列には src_id と dest_id の関連を示すキーワードを格納することによってデータ間の関係を保持する。表 1 の第一

**表1**　関連付け情報を格納するテーブル

| src_id | dest_id | link |
|---|---|---|
| 10080 | 10089 | phoneme |
| 10080 | 10088 | info |
| 10080 | 10085 | label |
| 10080 | 10086 | amp |
| 10080 | 10087 | fft |

行では，GuiAudio によるデータ送信によって，元の音声データの ID である10080が src_id 列に，音素列データの ID である10089が dest_id 列に入り，link 列に phoneme というキーワードが入ることによって，ID が10080のデータと ID が10089のデータが phoneme という関係で保持されたことを示している。音素以外のデータについても，元のデータとの関連が link 列の内容によって保持されている。システム内ではこの情報を基に，図 6 に示したような web ページを生成しており，データ間の関連を含むデータ送信の成功が確認できた。

## 8．ラベリングの試行実験

本研究で開発したプログラムによる音素ラベリングの結果を確認するために，音素ラベルが提供されている市販の音声データベースを用いて，本プログラムの開発者が音素ラベリングを試みた。使用した音声データベースは電気通信基礎技術研究所（ATR）の日本語音声データベース[9]であり，話者 MAU の音声データ 3 単語（/aikawarazu/, /aikyo:/, /aisatsu/）の延べ20音素のラベリングを行った。

データベース付属のラベルと，開発したプログラムによるラベリング結果の差は，20音素全ての平均で19ミリ秒であった。音素レベルでのラベリングについては，音素間の過渡部における境界決定が難しく，音素の境界規則によって結果が大きく異なる。またラベラーの熟練度の違いを考慮すると妥当なラベリング結果が得られたと考えられる。

## 9．まとめ

本研究では，Web サーバから音声データを取得し，GUI による操作で音素ラベリングを行うプログラムを開発した。また，音素ラベリングの結果と各種情報を即時にサーバに送信し，それらと元の音声データを関連付けてサーバに格納することができた。音声データと，それに付随する各種データは膨大な量となるため，管理が煩雑にならないよう即時に関連付けて保存できる意義は大きい。また，このプログラムを複数のコンピュータで利用すれば，手作業が避けられないラベリングを分担して行うことが可能であり，作業の効率化を図ることができる。

本研究で開発したアプリケーションプログラムをさらに発展させる手法としては，プラグイン等の形式で Web ブラウザに機能を組み込むことが考えられる。これにより，音声分析やラベリングの作業をブラウザのみで完結させることが可能になる。

## 参考文献

[1] 青木直史，伊藤博之，澤田 周，須藤健次，“XML による音声データベースのオープンコンテンツ化，”電子情報通信学会総合大会講演論文集，情報・システム(1)，no.SD-1-1， pp.”S-1”-”S-2”, March 2004.

[2] 今野英明，金光秀雄，高橋伸幸，外山淳，新保勝，“Web データベース汎用システムの開発と音声データ管理システムとしての利用，”電子情報通信学会技術報告， no.DE2005-16, pp.19-24, 2005.

[3] 小杉風友，今野英明，“音声分析を目的とした GUI クライアントアプリケーションの試作，”日本産業技術教育学会北海道支部研究論文集， no.24， pp.19-22, 2011.

[4] 小杉風友，今野英明，金光秀雄，“音声データの効率的管理を目的とした WebDB システムの開発—GUI クライアントの試作—，情報処理北海道シンポジウム2011講演論文集， no.C-5, pp.75-76, 2011.

[5] 今野英明，小杉風友，金光秀雄，高橋伸幸，“Web ベースデータ管理システムを用いた音声分析，”日本音響学会講演論文集， no.1-R-15, pp.463-464, March 2012.

[6] R. Nisimura, J. Miyake, H. Kawahara, and T. Irino, “Development of Speech Input Method for Interactive

VoiceWeb Systems," in Human-Computer Interaction, Part II, Lecture Notes in Computer Science, Volume 5611, ed. J.A. Jacko, pp.710-719, Springer-Verlag, Berlin, 2009.
[7] 青木直史，ブルガー アレキサンダー，山本 強，青木由直，“XML による音声データベースの構築とクライアント/サーバー音声合成システムの開発，”電子情報通信学会技術報告， no.DE2001-112， pp.33-38, 2001.
[8] E.R. Harold, Java ネットワークプログラミング，オライリー・ジャパン，東京，2001.
[9] 武田一哉，匂坂芳典，片桐滋，桑原尚夫，“音声データベース構築のための視察に基づく音韻ラベリング，”ATR Technical Report, no.TR-I-0019，1988.

（小杉　風友　株式会社エイチ・アイ・ディ）
（今野　英明　函館校准教授）
（金光　秀雄　函館校教授）